

BIGDATA Storage with Hadoop (HDFS Command Line)

I) Practice of HDFS Commands

1. In order to work with HDFS you need to use the **hadoop fs** command. For example to list the / and /user directories you need to input the following commands:

```
> hadoop fs -ls /  
> hadoop fs -ls /user
```

To do a recursive listing we'll use the **-lsr** command rather than just **-ls**. Try this:

```
> hadoop fs -lsr /user
```

For a file, it returns stat on the file with the format:

filename	<number of replicas>	size	modification_date	modification_time
permissions	userid	groupid		

For a directory, it returns list of its direct children as in UNIX, with the format:

dirname	<dir>	modification_date	modification_time	permissions	userid
groupid					

2. To make the directory *test* under */user/cloudera* you can issue the following command:

```
> hadoop fs -mkdir /user/cloudera/test
```

Now let's see the directory we've created:

```
> hadoop fs -ls /user/cloudera  
> hadoop fs -ls /user/cloudera/test
```

3. You should be aware that you can pipe (using the | character) any HDFS command to be used with the Linux shell. For example, you can easily use *grep* with HDFS by doing the following:

```
> hadoop fs -mkdir /user/cloudera/test2  
> hadoop fs -ls /user/cloudera | grep test
```

4. In order to move files from your regular linux filesystem to HDFS you will likely use the *put/copyFromLocal* command.

```
> hadoop fs -put /home/xxx /user/cloudera/xxx
```

Or

BIGDATA Storage with Hadoop (HDFS Command Line)

```
> hadoop fs -copyFromLocal /home/xxx  
/user/cloudera/xxx
```

You should now see a new file called /user/cloudera/xxx listed. In order to view the contents of this file we will use the `-cat` command as follows:

```
> hadoop fs -cat /user/cloudera/xxx
```

You should see the output of the xxx file (that is stored in HDFS).

5. In order to find the size of files you need to use the `-du` or `-dus` commands. Keep in mind that these commands return the file size in bytes. To find the size of the /user/cloudera/xxx file use the following command:

```
> hadoop fs -du /user/cloudera/xxx
```

To find the size of all files individually in the /user/cloudera directory use the following command:

```
> hadoop fs -du /user/cloudera
```

To find the size of all files in total of the /user/cloudera directory use the following command:

```
> hadoop fs -dus /user/cloudera
```

6. If you would like to get more information about a given command, invoke `-help` as follows:

```
> hadoop fs -help
```

For example, to get help on the `dus` command you'd do the following:

```
> hadoop fs -help dus
```

7. To copy files to the local file system from hdfs, use either `get` or `copyToLocal`.

Example:

```
> hadoop fs -copyToLocal /user/cloudera/xxx /home/xxx  
or  
hadoop fs -get /user/cloudera/xxx /home/xxx
```

To copy files from source to destination in hdfs

BIGDATA Storage with Hadoop (HDFS Command Line)

```
> hadoop fs -cp /user/cloudera/file1  
/user/cloudera/file2
```

8. To delete files use the following command:

```
> hadoop dfs -rm /user/cloudera/file1
```

To delete the files recursively, use the following command:

```
> hadoop dfs -rmr /user/cloudera/xxx
```

To displays last kilobyte of the file to stdout(Similar to UNIX tail command):

```
> hadoop dfs -tail /user/cloudera/file1
```

II) Merging files in a directory

Create a directory named 'random' under /user/cloudera. Copy two distinct files under that directory from local file system. How do you combine all the files into single one using hdfs commands?